Decomposing complementizers: the fseq of French, Modern Greek, Serbo-Croatian and Bulgarian complementizers

Baunaz, Lena

Abstract: This chapter discusses the morphosyntax of French, Modern Greek, Serbo-Croatian, and Bulgarian complementizers equivalent to English that. From long-distance wh-extractions across complementizers in these languages, it is shown that (i) the morpheme complementizer is composed of features that are hierarchically ordered according to a functional sequence (fseq) (see Baunaz 2015, 2016a; Baunaz and Lander to appear); (ii) the complementizer morpheme lexicalizes structures of different sizes; (iii) the distribution of complementizers is governed by veridicality (see Baunaz 2015, 2016a); (iv) the complementizer morpheme is syntactically active. The basic template for complementizers that I argue for is F4 > F3 > F2 > F1. Evidence in favor of this template comes from crosslinguistic patterns of syncretism and featural Relativized Minimality (Starke 2001; Rizzi 2004; Haegeman 2010, among others). Evidence in favor of different realizations of the complementizer is provided by means of long-distance extractions across declarative embedded clauses.

DOI: https://doi.org/10.1093/oso/9780190876746.003.0006

<Chapter 6 Decomposing Complementizers: The fseq of French, Modern Greek, Serbo-Croatian, and Bulgarian Complementizers♣>

Lena Baunaz

<1. Introduction>

It has been reported in the literature (Roussou 2010, Baunaz 2015, 2016a, among others) that complementizers vary crosslinguistically as to what information they lexicalize. This variation is illustrated in Table 6.1: French (Fr) *que*, Serbo-Croatian (SC) *da* and Bulgarian (Bg) *če* do not have a single analogue in Modern Greek (MG), but correspond to a number of different instantiations: *pu* introduces *epistemic factive* 'remember'-type complements (1a), and *oti* introduces *non-factive* 'say'-type complements (2a), whereas French (1b, 2b), SC (1c, 2c) and Bulgarian (1d, 2d) display the same complementizer – *que/da/če* – in these two contexts. As a first approximation, we conclude that French *que*, SC *da*, and Bulgarian *če* unite properties of multiple items in MG. This pattern (and similar ones noted in the literature) raises the question of how these different properties get spelled out as morphemes crosslinguistically.

Table 6.1 Declarative complementizers in MG, French, SC, and Bulgarian 1.0

	Factive verbs	*Non-factive verbs*
MG	pu	oti
Fr	que	
SC	da	
Bg	če	

(1) a. Thimame **pu** se sinandisa stin aghora. (MG)
remember.1SG that you met.1SG at.the market
(Roussou 1992, 125, (3))

b. Je me rappelle **que** je t'ai
I 1SG.ACC remember that I you.have
rencontré au marché (Fr)
meet.PAST.PART at.the market

c. Sjećam se **da** sam te upoznao
remember1.SG that AUX.PAST.1.SG you meet.PAST.PART
na tržnici. (SC)
on the.market

d. Pomnja, **če** te sreštnax na pazara (Bg)
remember1.SG that you meet.PAST.PART on the.market
'I remember that I met you at the market.'

(2) a. O Pavlos ipe **oti** i Roxani efije. (MG)
the Paul said.3SG that the Roxanne left.3SG
'Paul said that Roxanne left.' (Giannakidou 2009, 1886, (7))

b. Paul a dit **qu**'il a vu Marie (Fr)
Paul has said that.he has seen Marie

c. Pavao je rekao **da** je
Paul AUX.PAST.3SG say.PAST.PART that AUX.PAST.3SG
vidio Mariju. (SC)
see.PAST.PART Mary

d. Pavel kaza, **če** e vidjal Mary (Bg)
Paul said that he saw Mary
'Paul said that he saw Mary'

The facts summed up in Table 6.1 might lead one to think that French, SC, and Bulgarian have two homophonous complementizers; for instance for French we could postulate two *que*s, *que*$_1$ and *que*$_2$, where *que*$_1$ appears under factive verbs and *que*$_2$ under non-factive verbs, each corresponding to the more specialized MG complementizers, i.e. *que*$_1$ would correspond to *pu* and *que*$_2$ to *oti*. Rather than analyzing these overlaps in terms of homophony I consider phonologically identical complementizers to be cases of syncretism (see Baunaz and Lander Chapter 1, Section 3.3.1).

Before we try to assess the feasibility of decomposing complementizers into articulated structures, we need to complete the provisional picture in Table 6.1. First, MG, SC, and Bulgarian factives can select two different complementizers. Semi-factives can optionally select *oti* in MG (see Roussou 2010 and references cited there) (3a). Emotive factives can only select *pu* (3b).

(3) a. Thimame **oti/pu** dhjavaze poli. (MG)
remember.1SG that read-3SG much
'I remember that he used to read a lot./I remember him reading a lot.'
(Roussou 2010, 590, (17))

b. O Pavlos lipate ***oti/pu** efije i Roxani. (MG)
the Paul is.sad.3SG that left.3SG the Roxanne
'Paul regrets that Roxanne left.' (Giannakidou 2015, (7))

In addition to *da*, SC exhibits a second complementizer, *što*, that exclusively appears under emotive factive verbs, verbs under which *da* is impossible (4a). In this context, (some) speakers from Croatia (henceforth Hr.) have the option of using *da* (Tomislav Sočanac, p.c.).

Note also that some (Serbian, henceforth Sr.) speakers can optionally select *što* under semi-factives (4b) (Boban Arsenijevic, p.c.).

(4) a. Žalim **što/%da** si povrijedio Ivana. (SC)
regret.1SG that AUX.PAST.2SG hurt.PAST.PART John
'I regret that you hurt John.'

b. Znam **da/%što** si bio u Gentu.
know.1SG that AUX.PAST.2SG been in Ghent
'I know that you've been to Ghent./I'm familiar with the fact that you've been to Ghent.'

Če is not the only declarative complementizer in Bulgarian: *deto* is used under emotive factive verbs (alternating with *če* in this context; see Krapova 2010) (5a). As with SC, there is also speaker variation in Bulgarian: for some speakers, semi-factive verbs can optionally select *deto* (5b).

(5) a. Naistina sǎžljavam, **deto/če** ne otedlix poveče vnimanie na postrojkata. (Bg)
'I really regret that I did not devote greater attention to the construction'
(Krapova 2010, 26, (56a))

b. Pomnja, %**deto/če** te sreštnax na pazara. (Bg)
'I remember that I met you at the market/meeting you at the market'

Second, MG, SC, and Bulgarian have a distinct mood particle marking the subjunctive mood under *desiderative* 'wish'-type verbs (6).[1]

(6) a. Thelo **na** fiji o Kostas. (MG)
want.1SG SUBJ leave.3SG the Kostas

b. Želim **da** Ivan ode. (SC)
want.1SG SUBJ John leave.3SG
'I want John to leave.'

c. Iskam **da** ostanat decata (Bg)
want.1SG that stay.3PL children
'I want the children to stay.' (Krapova 1998, 86, (24b))

In all these contexts, French predicates select *que* throughout (7). In addition subjunctive mood is expressed via verbal morphology, as seen in (7b).

(7) a. Je me rappelle **que** Marie est partie. (Fr)
I 1SG.ACC remember that M. is.IND left
'I remember that Mary left.'

b. Je regrette/veux **que** Marie part-e.
I regret/want that M. leave-SUBJ
'I regret that Mary leaves/I want Mary to leave.'

In view of these observations, the table in Table 6.1 can be modified as in Table 6.2, where three columns have been added (emotive factive/semi-factive/desideratives), and more language-internal variations are considered (indicated by the % diacritic, *plus* the variety in question).

Table 6.2 Declarative complementizers in French, SC, Bulgarian, and MG, 2.0

	emotive factive	*semi*-factive	*non*-factive	desiderative
Fr	que	que	que	que
SC	% Hr. da	da	da	da
	što	% Sr. što		
Bg	če	če	če	da
	deto	% deto		
MG	pu	pu	oti	na
		oti		

<1.1 Proposals>

In this paper, I make four major claims: (i) the complementizer morpheme is composed of features that are hierarchically ordered according to a functional sequence (fseq) (see Baunaz 2015, 2016a; Baunaz and Lander to appear)[2]; (ii) the complementizer morpheme lexicalizes structures of different sizes; (iii) the distribution of complementizers is governed by veridicality (see Baunaz 2015, 2016a); (iv) the complementizer morpheme is syntactically active. The basic template for the internal structure of complementizers is that in (8).

(8) $F_4 > F_3 > F_2 > F_1$

Evidence in favor of (8) comes from crosslinguistic patterns of syncretism and Relativized Minimality (RM) (Starke 2001; Rizzi 2004).

<1.2 Syncretism and Relativized Minimality>

The nanosyntactic approach to syncretism (Caha 2009) is based on the idea that features are additive. As a consequence, an fseq can realize several structures (see Baunaz and Lander Chapter 1, [INSERT PAGES], Lander and Haegeman Chapter 5, [INSERT PAGES]). The structures in (9) can each be matched to a phonological form. So /na/ matches F_1P in (9a),

/oti/ matches the combination of F_1 with F_2 (in this order), namely F_2P in (9b), etc. We speak of syncretism when two (or more) structures are associated to the same phonological form, as in (9c) and (9d). A syncretism thus represents a one-to-many mapping between phonological form and syntactic structure.

(9) a. $[_{F1P}\ F_1\ ]$ => /na/

b. $[_{F2P}\ F_2\ [_{F1P}\ F_1]]$ => /oti/

c. $[_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1\ ]]]$ => /pu/

d. $[_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1\ ]]]$ => /pu/

As has become clear from Chapter 1 (Sections 3.3.1 and 3.3.2, and references cited there) what has become known as the *ABA generalization plays a crucial role in the study and interpretation of syncretism and for the identification of the features and their relative position in the fseq. It states that syncretism only targets contiguous structural layers, i.e. ABA patterns are ruled out. Consider Table 6.3. Table 6.3 has been stripped from its Serbian vs. Croatian and Bulgarian variations (i.e. % in Table 6.2) and the columns have been relabeled as in (9), for simplicity's sake. See below for a more thorough account.

Table 6.3 Declarative complementizers in French, SC, Bulgarian, and MG (simplified version)

	F_4	F_3	F_2	F_1
Fr	que	que	que	que
SC	što	da	da	da
Bg	deto	če	če	da
MG	pu	pu	oti	na

The syncretisms displayed by French, SC, Bulgarian, and MG in Table 6.3 suggest that the adjacency relations are as displayed in (10).

(10) $F_4 \mid F_3 \mid F_2 \mid F_1$

The linear ordering in (10) is the only one which can capture these facts without any *ABA patterns in MG, SC, and Bulgarian (and French): *što*, *deto*, and *oti* do not intervene between the different realizations of *da*, *če*, and *pu*. Bulgarian shows that F_3 and F_2 must be adjacent, and MG that F_4 and F_3 are contiguous.

What the observed syncretisms and the *ABA theorem leave open is which *hierarchical order* in (11) is appropriate.

(11) a. $F_4 > F_3 > F_2 > F_1$

b. $F_1 > F_2 > F_3 > F_4$

The strategies I adopt here to detect the fine-grained structure of complementizers are based on (i) syncretism (see Chapter 1, Section 3.3.1) and (ii) a development of featural RM (Starke 2001; Rizzi 2004, 2013; Haegeman 2010; Haegeman and Ürögdi 2010ab; among others). In particular, (ii) is based on the following hypothesis: the less semantically marked an item is, the less structure it has (Starke 2001). In other words, the item with the smallest structure is semantically unmarked. Depending on which end of the $F_4 \mid F_3 \mid F_2 \mid F_1$ spectrum the item with less semantics is syncretic with, we should be able to identify the 'smaller' end of the hierarchy.

<2. Strong vs. Weak presupposition>

In Section 1, we have seen that MG semi-factives can select two different complementizers, and so do SC/Bulgarian semi-factives and emotive factives.

In MG semi-factives may indeed select *oti* (see Table 6.2), but they do so only when they involve *weak presupposition*; when they express strong presupposition they select *pu*, see Roussou (2010, and references cited there). The use of *oti* in (12) somehow indicates some

factivity weakening of the main predicate. The use of *pu*, however, commits the speaker to the truth of the embedded proposition. In (12), the continuation *but he is wrong, because I didn't* forces a reading where the speaker's point of view about the truth of the embedded proposition needs to be taken into account. In the context in (12), only the *oti*-version is felicitous.

(12) a. O Janis paraponethike **oti** ton ksexasa; ala
the John complained.3SG that him forgot.1SG but
kani lathos: dhen ton ksexasa.
make.3SG mistake not him forgot.1SG

b. O Janis paraponethike **pu** ton ksexasa; # ala
the John complained.3SG that him forgot.1SG but
kani lathos: dhen ton ksexasa
make.3SG mistake not him forgot.1SG
'John complained that I forgot him; but he is wrong, because I didn't.'

(Giannakidou 2015, (92))

Baunaz (2016a) claims that in situations like (12), *oti*-selection involves a CP which expresses truth commitment by the subject exclusively (vs. speaker), i.e. the truth of the embedded proposition is relative to the speaker. (Some) MG semi-factives are then able to shift their meaning from strong to weak presupposition, i.e. these verbs are no longer semi-factive, but range over emotive factives.

As shown in Section 2 (see (4) above), semi-factives embed *da* in SC. Some Serbian speakers can also use *što* in this context. When *da* is embedded, the truth value of the embedded proposition must be true; when *što* is embedded, the factivity presupposition is

weakened. Emotive factives embed *što* in SC. Some Croatian speakers can also use *da* in this context.[3] The choice of complementizer coincides with presupposition strength: *da* involves *weak presupposition* and *što* involves *strong presupposition*. In case *što* is selected with semi-factives, or *da* with emotive factives, factivity is weakened (Boban Arsenijevic, Tomislav Sočanac, p.c.). It is important to note that when the verb triggers weak presupposition, factivity is not canceled out, as the embedded sentence needs to be considered true in the discourse: "the speaker speaks on the condition that it's true, but does not commit to it. The truth is based on the information contributed by someone else (the hearer most typically), not the speaker." (Baunaz 2016a, 71; Boban Arsenijevic, p.c.), i.e. the truth of the embedded proposition is relative to the speaker.

Bulgarian is very similar to SC, in that respect: recall that Bulgarian has two declarative complementizers: *deto* and *če*. Some (factive) verbs appear to optionally select both. For instance, Krapova (2010) reports that a (sub-)type of emotive factives can select *deto* as well as *če*. Some (other) speakers also optionally accept *deto* with semi-factives (especially 'remember') (Teodora Radeva-Bork, p.c.). As is the case for Serbian and Croatian *što* vs. *da* alternations, the choice of *če* vs. *deto* makes a difference for both types of speakers: *če* involves a weak presupposition, *deto* a strong presupposition.

Our findings have been summed up as in Table 6.2, repeated here as Table 6.4 for conveninence.

Table 6.4 Declarative complementizers in French, SC, Bulgarian, and MG, 2.0

	emotive factive	*semi*-factive	*non*-factive	desiderative
Fr	que	que	que	que
SC	% Hr. da	da	da	da
	što	% Sr. što		
Bg	če	če	če	da
	deto	% deto		
MG	pu	pu	oti	na
		oti		

As it stands, the presentation in Table 6.4 does not allow us to make a distinction between weak vs. strong presuppositions in the factive columns. What SC, Bulgarian, and MG tell us is that (some) semi-factives can shift their meaning to a weak presupposition interpretation (i.e. these verbs are no longer semi-factive, but must be ranged as emotive factive instead), while some emotive factives may shift their meaning to a strong presupposition interpretation (i.e. these verbs are no longer emotive factive, but semi-factive), i.e. verbs can switch classes. The form of the complementizer indicates this switch.

The weak vs. strong presupposition discussed here recalls the distinction *relative* vs. *strong veridicality* discussed in Baunaz and Puskás (2014) for French. I propose that the notion of veridicality can handle the apparent dual nature of these verbs insightfully. In other words, Table 6.4 can be refined in light of the notion of veridicality. This will require our columns to be relabeled.

<3. Veridicality>

On the basis of French, Baunaz and Puskás (2014) have argued that the classification of verbs involved in the selection of the complementizers in Table 6.4 should be refined in terms of veridicality, rather than factivity. Giannakidou (2009) defines the notion of veridicality as in (13), i.e. an embedded proposition has to be true for at least one individual (the subject of the main verb and/or the speaker), in all the worlds of a relevant model.

(13) **Veridicality** (Giannakidou 2009, 1889)

a propositional operator *F* is veridical iff from the truth of *Fp* we can infer that *p* is true according to some individual *x* (i.e. in some individual *x*'s epistemic model)

Baunaz and Puskás (2014) investigate the notion of 'some individual' in (13) to understand how it applies to emotive factive, semi-factive, and non-factive complements. They show that with semi-factives, the embedded proposition must be true, both from the subject's and the speaker's point of view, as shown by the continuation in (14).

(14) Paul découvre que Marie est partie, # mais c'est faux (Marie n'est pas partie).
'Paul found out that Marie left, but it is false (Marie didn't leave)'

With emotive factives, the embedded proposition must be true from the point of view of the subject, but not (necessarily) from that of the speaker, as shown by the continuation in (15).

(15) Jean est persuadée qu'il pleut, et il **regrette** qu'il pleuve *(but of course it's not raining).*
'Jean is convinced that it's raining, and he regrets that it's raining.'

(from Schlenker 2005, 27, fn.14, (i))

Baunaz and Puskás (2014, 245) "observe that the shift in the relevant epistemic model (i.e. of the Speaker or of the Subject) allows to make different inferences with respect to the truth of the embedded proposition, i.e. the veridical status of *regretter* 'regret' (…) is relative", as opposed to that of semi-factives.

Finally negating the complements of non-factives (verbs of saying, desideratives) does not yield contradictory statements, i.e. these verbs do not infer the truth of their complement, by neither the subject or the speaker (16).

(16) She is convinced that it's raining, and...

a. **veut/suggère** qu'il pleuve. *(But of course it's not raining!)*
wants/suggests that it rain.SUBJ

b. **dit** qu'il pleut. *(But of course it's not raining!)*
says that it rains.IND

They also observe that there are veridical verbs in French that can switch from one type of veridicality to another, as illustrated in (17). See also Baunaz (2016a).

(17) Pierre wrongly believes that Marie is getting married, and...

a. # réalise/comprend qu'elle ne **veut** pas rester célibataire.
'realizes/understands that she doesn't want.IND to stay single'

b. regrette/comprend qu'elle ne **veuille** pas rester pas célibataire.
'regrets/understands that she doesn't want.SUBJ to stay single'

The predicates in (17) are all veridical in Giannakidou's sense, as both (17a) and (17b) are true according to some individual *x*, namely Pierre and/or the speaker. In addition to the subject, *p* must also be true for the speaker, as the continuation shows in (17a). In contrast, if *regrette/comprend* in (17b) presupposes that *x* believes that *p* : *p* is true, but only with regard to the epistemic model of Pierre (and not (necessarily) with regard to that of the speaker) (Schlenker 2005). So "the shift in the relevant epistemic model (that is, of the Speaker or of

the Subject) [permits] to make different inferences with respect to the truth of the embedded proposition" Baunaz and Puskás (2014, 246). The two uses of *comprendre* involve different features (strongly veridical vs. relatively veridical). In that sense there must be different verbal realizations of the same phonological form (= syncretism).

To sum up, the distinction weak vs. strong presupposition is similar to the distinction *relatively* vs. *strongly* veridical: strong presuppositional verbs are **strongly veridical**, in that they require that the embedded proposition be true from the point of view of the speaker and from that of the subject; (ii) weak presuppositional verbs are **relatively veridical** in that they require that the embedded proposition be true from the point of view of the subject, but not (necessarily) from that of the speaker (cf. Schlenker 2005; see also Giannakidou 1998 for MG and Tóth 2008 for Hungarian); (iii) verbs of saying and desideratives do not embed propositions whose truth must be inferred by the subject or by the speaker: they are **non-veridical**.

I claim that these distinctions also apply to the seemingly ambiguous verbs discussed in Section 2. Semi-factives can have two readings in MG and some Bulgarian/SC: strongly and relatively veridical; emotive factives can also involve two readings in SC and for all Bulgarian: strongly and relatively veridical. Just like for French *comprendre* in (17), I claim that because the verbal realizations allowing shift of meaning have the same phonological form, they must be syncretic items. Crucially this meaning shift correlates with the form of the complementizer in Bulgarian (*deto* vs. *če*), suggesting that it is veridicality that selects the type of complementizer, as seen in Table 6.5.[4]

Table 6.5 Relatively/strongly veridical complementizers in MG, Bulgarian, SC, and French

	relatively veridical	*strongly veridical*
MG	pu	pu
	oti	
Bg	če	deto
SC	% Hr. da	da
	% Sr. što	što
Fr	que	que

The syncretism patterns of MG and Bulgarian tell us that the relatively and strongly veridical columns should be inverted: the relatively veridical and non-veridical columns must be contiguous (see MG *oti* and Bulgarian *če*), to avoid giving rise to an ABA violation. The syncretism patterns of SC and MG also tells us that the strongly veridical and relatively veridical columns should be adjacent too (see SC *da*, *što* and MG *pu*). In addition, note that the non-veridical column should be split into two columns, one hosting complementizers selected by verbs of saying (taking the indicative mood), and another selected by desiderative verbs (taking the subjunctive mood in MG, Bulgarian, SC, and French; see Sočanac (2017) for SC). Non-veridical verbs taking the indicative mood select *če* in Bulgarian, while non-veridical verbs selecting the subjunctive mood select *da*. *Če* being syncretic with relatively veridical complementizers, the two columns should be adjacent. This gives us Table 6.6, and the linear ordering in (18):

Table 6.6 Declarative complementizers MG, Bulgarian, SC, and French, 3.0.

	strongly veridical	*relatively veridical*	non-veridical$_{IND}$	non-veridical$_{SUBJ}$
MG	pu	pu	oti	na
		oti		
Bg	deto	če	če	da
SC	da	% Hr. da	da	da
	što	% Sr. što		
Fr	que	que	que	que

(18) strongly veridical | relatively veridical | non-veridical$_{IND}$ | non-veridical$_{SUBJ}$

From now on, I call the complementizers selected by strongly/relatively and non-veridical verbs *strongly/relatively* and *non-veridical* complementizers, respectively.

The syncretism patterns do not give us any insight into the ultimate hierarchical relation of the features in (18), though: it only helps us deduce a linear order (see Section 1.2). In Section 4 I argue that non-veridical complementizers occupy the right edge of our fseq, and that strongly veridical complementizers its left edge. My proposal is based on featural RM.

<4. Veridical islands (formerly known as *factive islands*)>

Deconstructing the meaning of *wh*-phrases escaping weak islands (WI), Starke (2001) observes that they are always interpreted with wide scope of existential presupposition (= what he calls *specificity,* β in (19-20)). He observes that when extraction out of WIs is tolerated, the moved *wh* must have something more than its intervener (which in that case, does not intervene at all), i.e. β in (19b); when extraction is blocked, it is either because the intervener has something more than the moved *wh*, i.e. it is bigger, since β is involved, as in (20a), or it is because the intervener and the *wh*-extractee share the same feature(s), i.e. they are of the same size (19a, 20b). When extraction is not blocked, it is because the intervener has something less than the moved *wh*, i.e. it is smaller, since β is *not* involved (19b) (the examples in (19) are from Haegeman and Ürögdi 2010a, 126, (23a-b)).

(19) a. * How do you wonder whether John will solve the problem? = *α_i ... α_j ...α_i

b. ? Which problem do you wonder whether John will solve? = $\alpha\beta$ ... α ... $\alpha\beta$

(20) a. ?? Who do you wonder which boy likes? = *α ... αβ ... α

b. ?* Which girls do you wonder which boy likes? = *αβ ... αβ ... αβ

In Starke's terms, this means that a quantifier α involving β, i.e. a specific *wh*-phrase like, for instance, *which* in (19b)/(20b), is semantically marked (it is 'specific'). The more semantically marked, the bigger (the *wh*-phrase contains α and β); the less semantically marked, the smaller (the *wh*-phrase only contains α).

Drawing on observations from syncretism patterns in Sections 2-3, we know that complementizers come in various sizes (cf. (18)), but we do not know the hierarchical ordering of their features. Featural RM can help us decide which featural hierarchy is correct when it comes to the internal structure of complementizers: 'bigger' complementizers will block movement of *wh*-phrases, 'smaller' complementizers will allow *wh*-phrases to move across them.

Since most of the verbs involved in Section 1 are generally claimed to induce weak (factive) islands in languages like English (Melvold 1991, among others), this section is thus based on long-distance *wh*-extraction across factive and non-factive domains. Recall that (non)-factivity should be considered as (non-)veridicality (Section 3). The relevant (non-)veridical selecting verbs are exemplified in Table 6.7 for MG, SC, Bulgarian, and French.

Table 6.7 Some (non-)veridical verbs in MG, SC, Bulgarian, and French.

	English translation	MG	SC	Bg	Fr
Strongly veridical	'remember' 'regret' 'understand'	*thimame*	*sjetiti se* %*žaliti*	*pomnja sâžaljavam*	*se rappeler* *comprendre*
Relatively veridical	'remember' 'regret' 'understand'	*thimame lipame*	%*sjetiti se žaliti*	*pomnja sâžaljavam*	*regretter* *comprendre*
Non-veridical	'say' 'want'	*leo thelo*	*reći željeti*	*kazvam iskam*	*dire vouloir*

I propose that the verbs in Table 6.7 select for different complementizers (creating different syntactic domains) of different 'sizes'. Thinking in terms of cumulative layers, this means that complementizers with the smallest feature composition have the least semantics, and that complementizers with the biggest feature composition have the most semantics. If that is correct, the size of complementizers should play a role in strong, weak, or no island configurations, just like the size of the moved element should also be relevant. In the remainder of this section, long-distance *wh*-extractions in MG, SC, Bulgarian, and French are closely studied in the configuration in (21), with the selecting contexts in Table 6.7.

(21) [argument $_{wh}$/adjunct$_{wh}$]…***Comp***…~~argument $_{wh}$/adjunct$_{wh}$~~

Sections 4.1-4.3 introduce extraction configurations in French, MG, SC, and Bulgarian. Section 4.4 discusses an apparent inconsistency in SC with long-distance extraction out of *što*-clauses.

<4.1 Long-distance extraction across veridical domains[5]>

In MG, long-distance *wh*-extractions across relatively veridical domains (with *pu* and *oti*) result in WI type violations (22-23). Long-distance extraction across strongly veridical domains (with *pu*) yields stronger violations, even when objects are involved ((22) vs. (24)).

(22) a. *Pjon*$_i$ thimase **oti** sinandises t_i? (MG)
who remember.2SG that met.2SG
'Who do you remember that you met?'

b. ?? *Pote*$_i$ thimase **oti** sinandises Maria t_i?
when remember.2SG that met.2SG Mary
'When do you remember that you met Mary?'

(23) a. *Pjon*$_i$ lipase **pu** pligoses t_i? (MG)
who be.sorry.2SG that hurt.2SG
'Who do you regret that you hurt?'

b. * *Pote*$_i$ lipose **pu** efijes t_i?
when be.sorry.2SG that left.2SG
'When do you regret that you left?' (Giannakidou 1998, 220, (70))

(24) a. * *Pjon*$_i$ thimase **pu** sinandises t_i? (MG)
who remember.2SG that met.2SG
'Who do you remember that you met?'

b. * *Pote*$_i$ thimase **pu** sinandises ti Maria t_i?
when remember.2SG that met.2SG the Maria
* 'When do you remember that you met Maria?' (Roussou 1992, 126, (7))

In SC spoken in Croatia, long-distance *wh*-extractions across the relatively veridical complementizer *da* is only possible with arguments, (25a), and is even preferred if *koga* 'who' is D-linked (Tomislav Sočanac, p.c.). Adjunct extraction is never possible (25b).[6] Long-distance extraction across the strongly veridical complementizer *da* in SC results in stronger restrictions (even with D-linked objects) (26).

(25) a. (?) *Koga*$_i$ žališ **da** si povrijedio t_i? (SC)
who regret.2SG that AUX.PAST.2SG hurt.PAST.PART
'Who do you regret that you hurt?'

b. * *Kad*$_i$ žališ **da** si otišao t_i?.
when regret.2SG that AUX.PAST.2SG leave.PAST.PART
'When do you regret that you left?'

(26) a. ?? *Koga*$_i$ se sjećas **da** si upoznao t_i? (SC)
who remember.2SG that AUX.PAST.2SG meet.PAST.PART
'Who do you remember that you met?'

b. * *Kad*$_i$ se sjećas **da** si upoznao Mariju t_i?
when remember.2SG that AUX.PAST.2SG meet.PAST.PART Maria
* 'When do you remember that you met Maria?'

Recall that in Bulgarian, the form of the complementizer distinguishes between relatively and strongly veridical domains: strongly veridical verbs select *deto*, relatively veridical verbs embed *če*. Relatively veridical *če* gives rise to WIs (27-28); when strongly veridical *deto* is used, extraction is (almost) never possible (or degraded, compared to extraction across *če* in similar contexts), no matter which type of verb is involved (29-30).

(27) a. *Koj*$_i$ pomniš, **če** sreštna na pazara t_i? (Bg)
who remember.2SG that met at market.the

b. * *Koga*$_i$ pomniš, **če** sreštna Maria na pazara t_i?
when remember.2SG that met Maria at market.the

(28) a. *Kakvo*$_i$ săzaljavaš **če** Ivan e napravil t_i? (Bg)
what regret.2SG that Ivan AUX.3SG done

b. * *Kăde*$_i$ săzaljavaš **če** Ivan e otišăl t_i?
where regret.2SG that Ivan AUX.3SG gone (Krapova 2010, (66))

(29) a.?(?) *Koj*$_i$ pomniš, **deto** sreštna na pazara t_i? (Bg)
who remember.2SG that met at market.the

b. * *Koga*$_i$ pomniš, **deto** sreštna Maria na pazara t_i?
When remember.2SG that met Maria at market.the

(30) a. * *Kakvo*$_i$ săzaljavaš **deto** Ivan e napravil t_i? (Bg)
what regret.2SG that Ivan AUX.3SG done

b. * *Kâde*$_i$ săzaljavaš **deto** Ivan e otišăl t_i?
where regret.2SG that Ivan AUX.3SG gone (Krapova 2010, (65))

In French only argument-extractions of D-linked *wh*-phrases are allowed with relatively veridical verbs (31-32). Aggressively non-D-linked *qui diable* 'who the hell' is rejected in (31b). With strongly veridical verbs, long-distance *wh*-extractions are never possible (33-34).

(31) a. *Quel tournoi* est-ce que Paul regrette **que** Roger ait vendu?

'Which tournament does Paul regret that Roger sold?'

b. * *Qui diable* est-ce que Paul regrette **que** Roger puisse aimer?

'Who the hell does Paul regret that Roger can love?'

c. * *Comment* est-ce que Paul regrette **que** Roger ait gagné le tournoi?

'How does Paul regret that Roger won the tournament?'

(32) a. *Quelle photo* est-ce que Jean comprend/admet **que** Jules prenne?

'Which picture does Jean understand/admit that Jules takes.SUBJ?'

b. ?? *Comment* est-ce que Jean comprend/admet **que** Jules prenne la photo?

'How does Jean understand/admit that Jules takes.SUBJ the picture?'

(Baunaz and Puskás 2014, 236 (6b), (7b))

(33) a. ??/* *Quelle photo* est-ce que Jean se rappelle **que** Jules prend?

'Which picture does John remember that Jules takes?'

b. * *Qui diable* est-ce que Jean se rappelle **que** Roger peut aimer ?

'Who the hell does John remember that Roger can love?'

c. * *Comment*$_{i}$ est-ce que Jean se rappelle **que** Jules prend la photo t_{i}?

*'How does John remember that Jules takes the picture?'

(34) a. * *Quelle photo* est-ce que Jean comprend/admet **que** Jules prend?

'Which picture does Jean understand/admit that Jules takes.IND?'

b. * *Comment* est-ce que Jean comprend/admet **que** Jules prend la photo?

'How does Jean understand/admit that Jules takes.IND the picture?'

(Baunaz and Puskás 2014, 235 (4b), (5b))

Wh-extraction out of veridical domains yields thus either strong or weak islands. In the next section, *wh*-extraction out of non-veridical domains is studied.

<4.2 Long-distance extraction across non-veridical domains>

Consider (35-42). In these examples, long-distance extraction across non-veridical domains (including verbs of communication and desideratives) is unproblematic, irrespective of the language and form of the complementizer, i.e. with both *oti* and *na* in MG (35-36), *da* in SC (37-38), *če* and *da* in Bulgarian (39-40) and *que* in French (41-42).

(35) a. *Pjon* ipe o Pavlos **oti** idhe? (MG)
who said.3SG the Paul that saw.3SG
'Who did Paul say that he saw?'

b. *Pote* ipes **oti** idhes ton Pavlo?
who said.2SG that saw.2SG the Paul
'When did you say that you saw Paul?' (Giannakidou 1998, 220, (69))

(36) a. *Pjon* theli o Pavlos **na** dhi? (MG)
who want.3SG the Paul SUBJ see.3SG
'Who does Paul want to see?'

b. *Pjon* theli o Pavlos **na** fiji?
who want.3SG the Paul SUBJ leave.3SG
'When does Paul want to leave?' (Giannakidou 1998, 220, (70))

(37) a. *Koga* je Pavao rekao **da**
who AUX.PAST.3SG Paul say.PAST.PART that
je vidio? (SC)
AUX.PAST.3SG see.PAST.PART
'Who did Paul say that he saw?'

b. *Kad* si rekao **da** si
when AUX.PAST.2SG say.PAST.PART that AUX.PAST.2SG
vidio Pavla?
see.PAST.PART Paul
'When did you say that you saw Paul?'

(38) a. *Koga* Ivan želi **da** vidi? (SC)
who Ivan want.3SG SUBJ see.3SG
'Who does John want to see?'

b. *Kada* Ivan želi **da** ode?
when Ivan want.3SG SUBJ leave.3SG
'When does Ivan want to leave?'

(39) a. *Kogo* kaza Pavel, **če** e vidjal? (Bg)
who said Paul that AUX.3SG see.PART
'Who did Paul say that he saw?'

b. ? *Koga* kaza, **če** si vidjal Pavel?
when said that AUX.2SG see.PART Paul
'When did you say that you saw Pavel?'

(40) a. *Koj* iskash **da** dojde. (Bg)

who want.2SG SUBJ come.3SG

'Who do you want to come?'

b. *Koga* iskash **da** dojda?

when want.2SG SUBJ come.1SG

'When do you want me to come?'

(41) a. *Quel tournoi* est-ce que Paul dit/observe **que** Roger a gagné? (Fr)

'Which tournament does Paul say/observe that Roger won?'

b. *Qui diable* est-ce que Paul dit/observe **que** Roger peut aimer?

'Who the hell does Paul say/observe that Roger can love?'

c. *Comment* est-ce que Paul dit/observe **que** Roger a gagné le tournoi?

'How does Paul say/claim that Roger won the tournament?'

(42) a. *Quelle raquette* est-ce que Paul préfère/désire/veut **que** Roger vende? (Fr)

'Which tennis racket does Paul prefer/desire that Roger sells? /

Which tennis racket does Paul want him to sell'

b. *Qui diable* est-ce que Paul préfère/ désire/veut **que** Roger aime ?

'Who the hell does Paul prefer/desire that Roger love? /

Who-the-hell does Paul want Roger to love?'

c. *Comment* est-ce que Paul préfère/désire/veut **que** Roger gagne le tournoi ?

'How does Paul prefer/desire that Roger win the tournament? /

How does Paul want Roger to win the tournament?'

<4.3 Summary>

Extraction facts tell us that there is something which uniformly blocks long distance *wh*-movements completely, partially, or not at all. Recall also that when WIs are created, argument-extraction is better with D-linked phrases. So the constraint must be somehow related to the type of main clause predicate involved. See Figure 6.1.

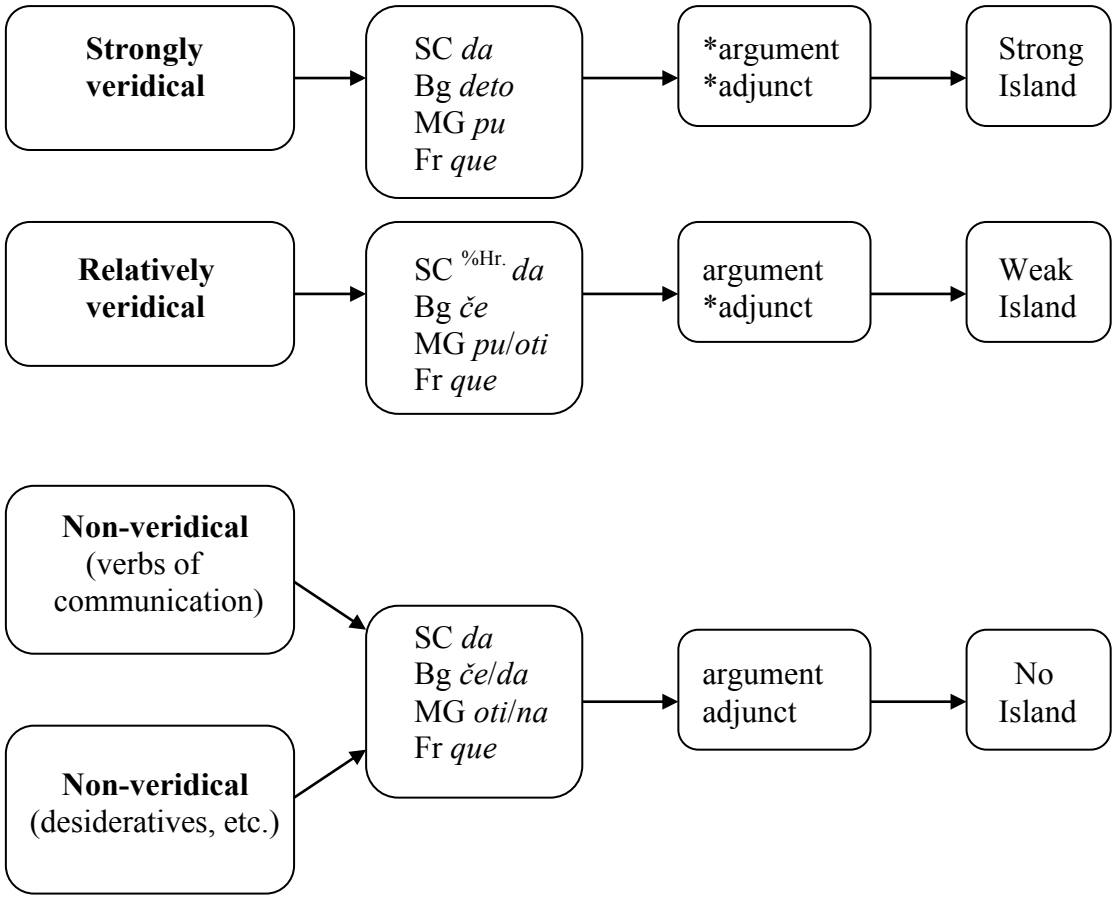


Figure 6.1

<4.4 Serbo-Croatian *što*>

For completeness' sake it should be added SC speakers can also embed *što* in strongly veridical domains and that Serbian *što* can also be selected under relatively veridical verbs (see Section 3, fn.4, as well as example (4) above, and Table 6.5 above). Arguments may be

extracted from strongly domains (43), but not from relatively veridical domains in SC spoken by some Serbians (44).

(43) a. (?) *Koga*$_i$ žališ **što** si *pro* povrijedio t_i? (SC)
who regret.2SG that AUX.PAST.2SG pro.ACC hurt.PAST.PART
'Who do you regret that you hurt?'

b. * *Kad*$_i$ žališ **što** si otišao t_i?
when regret.2SG that AUX.PAST.2SG leave.PAST.PART
'When do you regret that you left?'

(44) a. * *Koga*$_i$ se sjećas **što** si *pro* upoznao t_i? (Serbian)
who remember.2SG that AUX.PAST.2SG pro.ACC meet.PAST.PART
'Who do you remember that you met?'

b. * *Kad*$_i$ se sjećas **što** si upoznao Mariju t_i?
when remember.2SG that AUX.PAST.2SG meet.PAST.PART Maria
* 'When do you remember that you met Maria?'

In view of the examples discussed in Sections 4.1-4.3, the judgments here are unexpected.

There are (at least) three reasons to think that *što*-clauses are not embedded clause structures. First, while *što*-clauses require resumption (Boban Arsenijevic, p.c.), as shown in (45) with resumptive *ga* 'it'. This requirement is rescued with a *pro* resumptive with nominative and animate accusatives (Boban Arsenijevic, p.c.). *Da*-clauses do not require resumption.

(45) Zaista žalim **što** sam *ga* se odrekao.
really regret.1SG that AUX.1SG it.Cl.GEN REFL renounced.M.SG
'I really regret that I renounced it.' (Boban Arsenijevic, p.c.)

Since (43a) and (44a) involve an animate accusative, both involve an accusative *pro* (in italics).

Second, relativization with *što* requires resumption, too (46) (Boban Arsenijevic, p.c.; Mitrović 2012), i.e. *što*-clauses are arguably relative clause constructions (see Arsenijevic 2009 for SC; see also Aboh 2005 for Gbe; Krapova 2010 for Bulgarian; and Haegeman and Ürögdi 2010ab for arguments in favor of factive complements as relatives).[7]

(46) Žena **što** *ju* je skavi čovek pozvao na večeru
woman that her AUX each man invited on dinner
zahvalila mu se. (SC)
thanked him REFL?
'The woman every man invited to dinner thanked him.' (Mitrović 2012, (15b))

Finally *što*-clauses can be embedded under veridical verbs. When they are embedded under *žaliti* 'regret', they introduce strongly veridical domains (see Section 3). If that is correct, we expect *što* to be embedded 'by default' under *the fact*, as in (47). This is not borne out: only *da* can appear under *činjenica* 'fact' (Boban Arsenijevic, p.c.).

(47) Zaista žalim činjenicu **da/*što** sam (SC)
really regret.1SG fact that AUX.1SG
ga povrijedio.
him.ACC.Cl hurt.M.SG
'I really regret the fact that I hurt him'

For these reasons, I leave *što*-clauses out of the discussion here.[8]

<4.5 The fseq of complementizers>

Section 3 told us that the linear ordering of the features constituting the fseq of complementizers is one where the relatively veridical and non-veridical features are adjacent, and where the strongly veridical and relatively veridical features are also adjacent (see (18), repeated here as (48)).

(48) strongly veridical | relatively veridical | non-veridical$_{IND}$ | non-veridical$_{SUBJ}$

What this section teaches us is that strongly veridical complementizers yield stronger islands than relatively veridical complementizers, and that non-veridical complementizers create no island at all. Taking into account featural RM and the idea that layers are cumulative (see Section 1.2), non-veridical complementizers, which are (semantically) unmarked (they are non-presuppositional), should be structurally very small; strongly veridical complementizers, which are the 'more semantically marked' of the complementizers investigated here (they involve strong presupposition), must be quite big. Relatively veridical complementizers, involving some kind of 'weak presupposition', are less semantically marked than strongly veridical complementizers, but more marked than non-veridical complementizers, and as such

must be of an in-between size, by this logic. Thinking in terms of cumulative layers, then, strongly veridical complementizers are the biggest of all complementizers and include all the other layers. As such they occur at the very left end of the complementizer-fseq, and the smallest, non-veridical complementizers should occur at the very right end of the complementizer-fseq (49). Relatively veridical complementizers are then sandwiched in between.

(49) strongly veridical > relatively veridical > non-veridical$_{IND}$ > non-veridical$_{SUBJ}$

The hierarchy in (49) is based on syncretism patterns with complementizers and on *wh*-extractions out of (non-)veridical domains in four different languages. If the hierarchy is adequate, the labels attributed to the features appear to be misleading, though: veridicality is a (semantic) property of predicates, not complementizers. In Section 5, I propose that (non-)veridical predicates can select three types of complementizers, which vary in terms of existential presupposition, i.e. complementizers can be specific, partitive or non-presuppositional. What governs what kind of feature(s) is selected is, I claim, (non-) veridicality.

<5. The features of *wh*-phrases and complementizers>

Under the logic used in Section 4, if a complementizer blocks *wh*-extractions, it must share (some) features with *wh*-pronouns, features relevant to RM. If it does not block movement, then features relevant to RM are not involved. Building on Baunaz (2015) and Baunaz (2016b), I claim here that the fseq of complementizers is basically similar to that of *wh*-phrases: both involve an operator that is built independently and that is prefixed to a BaseP, as schematized in Figure 6.2.[9]

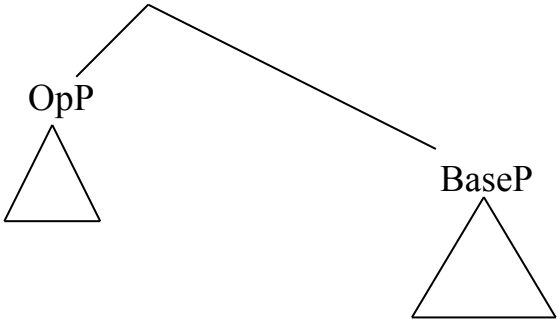


Figure 6.2

<5.1 Prefixed OpP>

As seen in Section 4, complementizers in MG, SC, Bulgarian, and French may block *wh*-movement, i.e. they may block movement of quantificational phrases. *Wh*-phrases minimally involve an Operator (Op) (Rizzi 1997, among others). So because they intervene in quantificational chains, (some) complementizers should also involve an operator (or an operator feature, see fn.9, and Haegeman and Ürögdi 2010 ab). I follow Roussou (2010), Manzini and Savoia (2011) and claim that complementizers may involve a **declarative** operator ($Op_{prop(ositional)}$, operating over propositional variables (in the embedded clause)). Unlike interrogative pronouns, they range over propositions, not individual variables.[10]

There is at least one reason to believe that Op is prefixed to complementizers: complementizers in French, MG, SC, and Bulgarian are syncretic with a morpheme that appears in the formation of quantifiers: Fr. *cha-**que*** 'every', MG *ká-**pu*** 'somewhere', Bg *kă-deto* 'where (relative)', etc. These quantifiers are bimorphemic and involve an overt Op morpheme that is prefixed to *-que/-pu/-deto*. Note that the MG complementizer *oti* is arguably also bimorphemic, consisting of the definite article *o-* plus *-ti* 'thing' (see Roussou 2010, among others) and conforms to the quantificational pattern above, see MG *ká-**ti*** 'something'.[11] Extending this idea to other quantificational elements amounts to saying that

these items are bimorphemic too, with Op being non-overt. In Figure 6.3, Op_{α} (α = wh, foc, $\forall$, $\exists$, etc.) is a constituent of its own, built in a domain outside of BaseP and subsequently prefixed to BaseP as independent element (see Baunaz and Lander Chapter 1, Section 3.3.5 and Starke Chapter 9; see also Baunaz 2016b; Baunaz and Lander to appear; and Leu's 2010 proposal on German *jeder* and Fr. *chaque*).

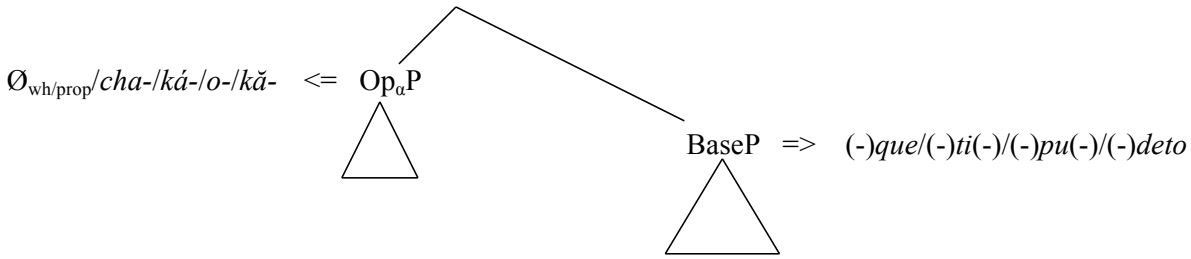


Figure 6.3

If OpP involves information about the semantic meaning of these constituents (namely interrogative, universal, etc.), what about BaseP? I propose in Section 5.2 that BaseP involves an fseq whose highest property spells out *specificity* (that is, β in Starke's terms discussed in Section 4). By the Superset Theorem, smaller pieces of the structure can also be spelled out, involving other features. These features, combined with Op_{α}, are what is relevant to RM and will account for the veridical islands presented in Section 4.

<5.2 BaseP>

Our working hypothesis, based on featural RM, is that *wh*-phrases and complementizers share similar features. In our structure in Figure 6.3, the features shared by *wh*-phrases and complementizers cannot be in PreP, as these features relate to the semantics of these items: *wh*-phrases are interrogative, complementizers are non-interrogative. Thus the feature we are looking for must rather belong to BaseP. In this section I show that several features may

appear in the extended projection of BaseP in Figure 6.3. On the basis of previous work of mine on *wh*-words in French I propose that the features making up BaseP are *specificity* and *partitivity*. I argue that these features are ordered above a categorial feature (*n(ominal)* for *wh*-phrases, c(ategory) for complementizers), which constitute the core of these constituents.

<5.2.1 Specificity, Partitivity, Neutrality>

Baunaz (2011, 2015, 2016b) assumes that *wh*-phrases can receive three different interpretations according to three different discursive contexts: *partitive*, *specific*, and *neutral.* These concepts are defined as follows: a **partitive** interrogative phrase is an object that belongs to a closed set of presupposed alternatives. A **specific** interrogative phrase narrows down the context to familiar individuals, excluding alternatives. As for *out-of-the-blue* interrogative phrases, they lack *specificity* and *partitivity*, i.e. they do not involve existential presupposition. In that sense, they are **neutral**.

(50) exemplifies partitivity, (51) specificity: the referent of *qui* in (50) can potentially be any girl belonging to the pre-defined set of girls, that is, the blonde one, the red one, or the brunette. In (51) the journalist wants to identify the individual that all the witnesses recognized.

(50) *After the parade, all the girls are standing in front of the jury. Joe, one of the judges, asks Bob*:

Joe: Et toi, **qui** tu préfères, là?

'And you, who do you prefer, here?'

Bob: La blonde. / La brune. / # Aucune.

'The blonde. / The brunette. / # None (of them).'

(51) *Witnesses and defendants have been brought face to face. One of the defendants has been accused by all the witnesses. Before the verdict, a journalist asks the judge:*

J: Et **qui** les témoins ont reconnu dans le box des accusés?

'Who is it that all the witnesses recognized in the defendants' box?'

Both partitive and specific contexts involve existential presupposition: in both (50) and (51), a negative answer is infelicitous, since it would go against the existential presupposition that there is an antecedent for *qui*. Importantly, these concepts are syntactically encoded: they play a role in Scope Islands (Baunaz 2011, 2016b). Starke (2001) shows that specificity plays a crucial role in *wh*-extraction out of WIs.

Interrogative phrases can also be used in contexts where the interlocutor does not know its referent (and if there is a referent available), that is in neutral context, as in (52).

(52) Stan : **Qui** as-tu croisé ce matin à l'entrainement ?

'Who did you meet this morning at practice?'

Nole : 'Nobody. / Roger.'

Qui can also appear in non-quantificational contexts, as is examplified in (53).

(53) a. Les professeurs rentraient chez eux, **qui** à Paris, **qui** à Bruxelles.

the professors returned home who to Paris who to Brussels

'The professors returned home, some of them to Paris, others to Brussels'

b. **Qui** apportait un fromage, **qui** un sac de noix, **qui** un quartier de chèvre…
who brought.3SG a cheese who a bag of nuts who a piece of goat…
'One brought a piece of cheese, one a bag of nuts, one a piece of goat meat...'
(Lipták 2001, 137, (13a))

In (53), *qui* is interpreted as a distributee. In (53a), *qui* refers to *professeurs* and is interpreted as partitive ('some of them'). In (53b), *qui* has no antecedent: it does not refer to a term that has been already mentioned, i.e. neither specificity nor partitivity is involved, and a neutral meaning shows up. It is an indefinite.

If we can show that one of these concepts is contained within the other, then we can argue that they are in a hierarchical relationship. The idea behind this is that the more semantically marked, the bigger, the less semantically marked, the smaller (cf. Starke 2001 and Section 4 above; see also Chapter 1, Section 3.3.1 about semantic composition as a tool to detect hierarchical structures). I propose that specificity contains partitivity, but not vice-versa. Swedish shows morphological evidence for this claim.[12]

<5.2.2 Morphological containment>

The semantics of the partitive and specific readings suggest a certain kind of containment. This can be seen in Swedish, for instance, where a partitive DP like *en av killarna* 'one of the boys' in (54a) may be preceded by a definite article to specify the reference of one of the individuals belonging to the presupposed set, i.e. turning the DP specific (54b):

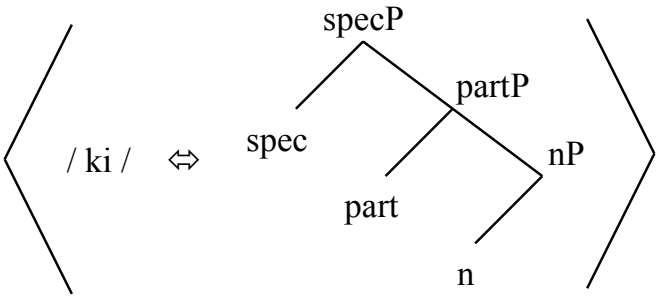
(54) a. [en av killarna] kom hem till mig igår (Swedish)
one of guys.the came home to me yesterday
'One of the guys came to my home yesterday.'

b. [den [en-a (av killarna)]], det vill säga, Eric…
the one-DEF of guys.the that wants say Eric…

What (54b) shows is that specificity is bigger than partitivity, and is as such more semantically marked. In specific contexts, partitive DPs can be used. This suggests that the two features are hierarchically ordered: specificity > partitivity.

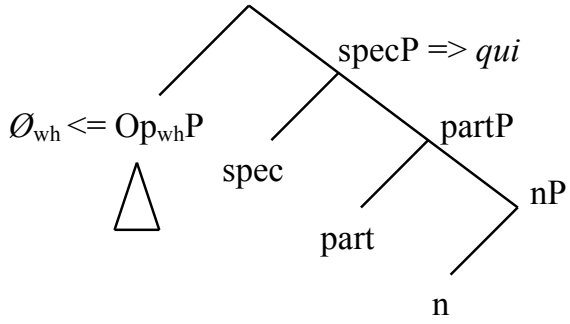
Qui thus displays three interpretations: specific, partitive, and neutral (recall that the operator part is built as an independent subtree and has by hypothesis nothing to do with our fseq; it is prefixed independently). These are syncretic in (50-53), i.e. there is a particular lexical tree (L-tree) for *qui* that maps onto a range of syntactic trees (S-trees) (Figure 6.4). The S-trees for *qui* show different feature make-ups and are of different sizes ((b) in Figure 6.4), with the partitive meaning contained within specificity. Non-interrogative *qui*, being non-quantificational, non-specific, and non-partitive is not prefixed with an OpP ((biv) in Figure 6.4). No special feature indicates the neutral reading, since the neutral reading is defined in term of the absence of both specificity and partitivity. nP indicates that *qui* is a (pro)nominal constituent (i.e. it belongs to a nominal fseq).[13]

a. Lexical tree for *qui*

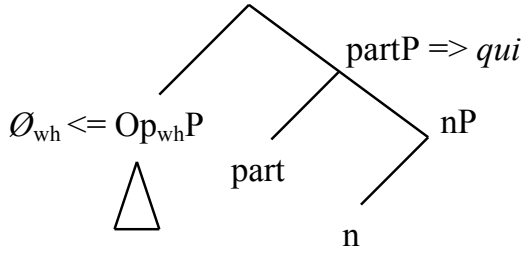


b. Syntactic trees for *qui* (with a prefixed operator)

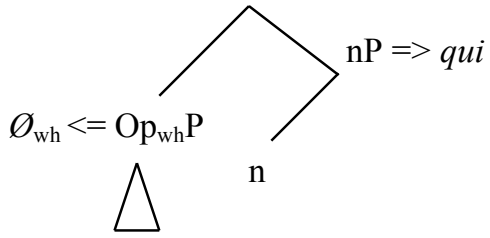
i. specific qui_{wh}



ii. partitive qui_{wh}



iii. neutral qui_{wh}



iv. neutral qui_{indef}

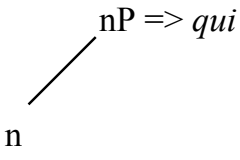


Figure 6.4

A reviewer notes that Figure 6.4 is reminiscent of Cardinaletti and Starke's 1999 classification of pronouns (see Chapter 1, Section 3.3.2). The reminiscence is due to the Superset Theorem, which is at the core of how syncretism is treated in nanosyntax. By the Superset Theorem, the L-tree in (a) in Figure 6.4 can match the S-trees in (b).

<5.3 The internal structure of complementizers>

From the beginning of this section our working hypothesis is that *wh*-phrases and complementizers share similar features, and the locus of this similarity is BaseP. Just like there exist specific, partitive, and neutral *qui*, I propose that complementizers may also be specific, partitive, and neutral. On a par with the fseq proposed for *qui*, I propose that complementizer *que* can lexicalize structures of different sizes. This analysis is extended to MG *pu/oti*, SC *da*, and Bg *če*/*deto.* I claim that veridicality governs the distribution of complementizers and selects different types of declarative complementizers: specific, partitive, or neutral, as in Figure 6.5 (c means here 'category', as Fr. *que* and MG *ti* are syncretic with nominal categories, while MG *pu*, Bg *deto* are syncretic with adverbials, and the category (or categories) of SC *da*, Bg *če* are left undefined here[14]).

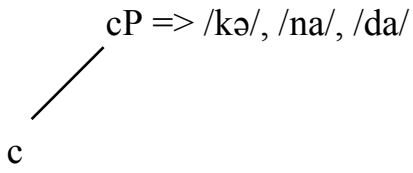
Non-veridical verbs select neutral complementizers*,* which range over non-finite sets of propositional variables (neither true, nor false). These complementizers are neither specific nor partitive. Also they are permeable to *wh*-arguments and *wh*-adjuncts alike, i.e. they do not involve $Op_{prop}P$. Because *wh*-phrases have a 'bigger' feature composition than these complementizers – *wh*-phrases at least involve $Op_{wh}P$ – they can move freely. Recall that non-veridical verbs can select (at least) two distinct types of complementizers (Section 2-3): complementizers selected under verbs of saying, and complementizers selected under

desideratives. So presumably there is no 'non-veridical' feature. The S-trees of these complementizers should, though, be different, even though the difference between them is minimal. Since that distinction is not related to veridicality, I refer to it as X here. Based on syncretism patterns of Bulgarian in Table 6.6, the complementizer selected by verbs of saying is syncretic with the one selected by relatively veridical verbs (= *če*), while for the complementizer selected by desideratives (*da*), I propose, based on RM (Section 4), that it lexicalizes the smallest bit of structure ((a) in Figure 6.5).

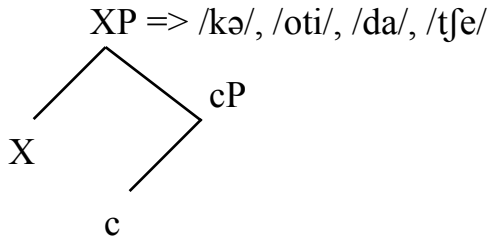
Complementizers selected by verbs of saying are slightly bigger and lexicalize X *plus* the c feature ((b) in Figure 6.5). Relatively veridical verbs select partitive complementizers which range over (a given set of) propositional variables (true or false). Partitive complementizers are quantificational: they create WIs. The consequence is that extraction is possible, but only for [specific] *wh*-arguments (which are 'bigger' than partitive complementizers (cf. (a) in Figure 6.4). *Wh*-adjuncts, which are non-specific and non-partitive, are not extractable (they are 'smaller' than partitive complementizers).

Strongly veridical verbs select specific complementizers that locate the complement proposition with respect to a given point of reference, binding a single propositional variable, which corresponds to a single truth value (true) (see Roussou 2010). Specific complementizers are quantificational: they block any type of *wh*-movement. In addition, they introduce *specific* propositions. Because specificity contains partitivity, the S-tree of specific complementizers contains a [specific] feature, which dominates a [partitive] feature, as in (d) in Figure 6.5. With specific complementizers, extraction is never possible: *wh*-/focus-phrases are blocked: they are either of the same size as specific complementizers, or smaller.

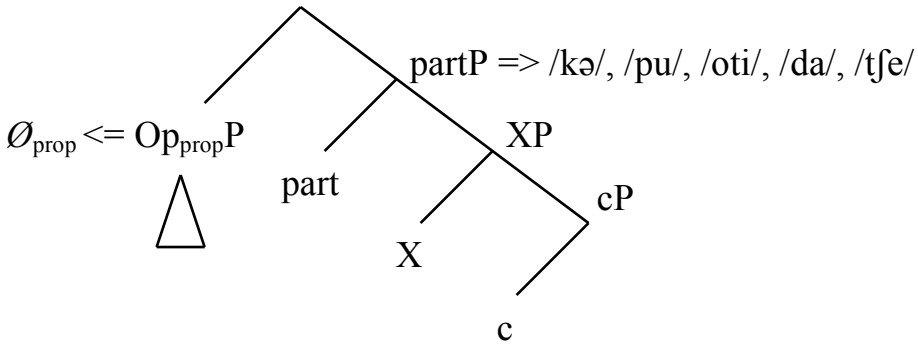
a. Bare complementizer



b. XP complementizer



c. Partitive complementizer



d. Specific complementizer

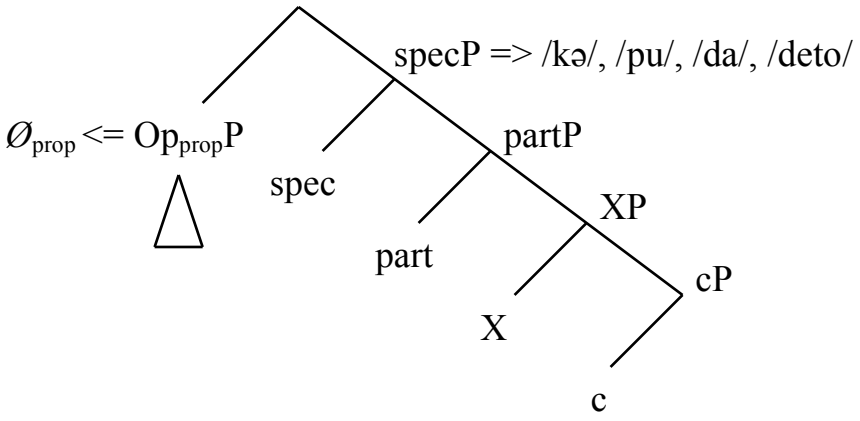


Figure 6.5

The size of complementizers is then relevant for intervention effects.

<6. Conclusion>

In this paper I have discussed the nanosyntax of French, MG, SC, and Bulgarian complementizers. Based on syncretism patterns and long-distance *wh*-extractions across complementizers in these languages, I have argued that (i) complementizers are complex morphemes, lexicalizing structures of different sizes; (ii) their distribution is governed by veridicality; (iii) complementizers are syntactically active. The intervention effects observed in the four languages have been accounted for in terms of (i) the type of features displayed by the relevant morphemes (complementizer, *wh*-phrase): i.e. OpP *plus* either the specific or the partitive feature is what matters to RM; (ii) the 'size' of the morphemes involved.

References

Aboh, Enoch. 2005. "Deriving Relative and Factive Clauses". In *Contributions to the 13th "IGG"*, edited by Laura Brugè, Giuliana Giusti, Nicola Munaro, Walter Schweikert and Giuseppina Turano. pp. 265-285. Venezia: Cafoscarina.

Arsenijevic, Boban. 2009. "Clausal Complementation as Relativization". *Lingua* 119: pp. 39-50.

Baunaz, Lena. 2011. *The Grammar of French Quantification* (Studies in Natural Language and Linguistic Theory 83). Dordrecht: Springer.

Baunaz, Lena. 2015. "On the Various Sizes of Complementizers". *Probus 27 (2)*: pp. 193-236.

Baunaz, Lena. 2016a. “Deconstructing Complementizers in Serbo-Croatian, Modern Greek and Bulgarian”. In *Proceedings of NELS 46* (1), edited by Christopher Hammerly and Brandon Prickett, pp. 69-77. Amherst, MA: Graduate Linguistics Student Association.

Baunaz, Lena. 2016b. “Wh-phrases in-situ: Interface Strategies”. *Acta Linguistica* Vol. 63 (2): pp. 125-168.

Baunaz, Lena and Genoveva Puskás. 2014. “The Selection of French Mood”. In *Variation within and across Romance Languages,* edited by Eric Mathieu and Marie-Hélène Côté, pp. 233-253. Amsterdam: John Benjamins.

Baunaz, Lena and Eric Lander. to appear. “Cross-categorial syncretism and the Slavic containment puzzle”. In *Balkan Syntax and (Universal) principles of Grammar,* edited by Iliana Krapova and Brian Joseph. Berlin: Mouton de Gruyter.

Baunaz, Lena and Eric Lander. (under review). “Deconstructing categories syncretic with the nominal complementizer”. Ms. University of Zurich and University of Gottenburg.

Caha, Pavel. 2009. *The Nanosyntax of Case*. Doctoral dissertation, Tromsø.

Cardinaletti, Anna and Michal Starke. 1999. “The Typology of Structural Deficiency. A Case Study of the Three Classes of Pronouns”. In *Clitics in the Languages of Europe*, edited by van Riemsdijk, Henk, pp. 145-233. Berlin: Mouton-De Gruyter.

Giannakidou, Anastasia. 1998. *Polarity Sensitivity as (Non)veridical Dependency*. Amsterdam/Philadelphia: John Benjamins.

Giannakidou, Anastasia. 2009. "The Dependency of the Subjunctive Revisited: Temporal Semantics and Polarity". [Special issue on *Mood*] *Lingua 119*: pp .883-1908.

Giannakidou, Anastasia. 2015. "Evaluative Subjunctive and Non-Veridicality". Ms. University of Illinois at Chicago.

Haegeman, Liliane. 2010. "The Internal Syntax of Adverbial Clauses". In *Exploring the left periphery*, edited by Kleanthes Grohmann and Ianthi Tsimpli. *Lingua thematic issue*, *120*: pp. 628- 648.

Haegeman, Liliane and Ürögdi, Barbara. 2010a. "Referential CPs and DPs: An operator movement account." *Theoretical Linguistics 36* : pp. 111-152.

Haegeman, Liliane and Ürögdi, Barbara. 2010b. "Operator movement, Referentiality and Intervention." *Theoretical Linguistics 36*: pp. 233-246.

Kayne, Richard S. 2009. "Antisymmetry and the Lexicon." *Linguistic Variation Yearbook 8* (1): pp. 1-31.

Krapova, Iliana. 1998. "Subjunctive Complements, Null Subjects and Case Checking in Bulgarian." *University of Venice Working Papers in Linguistics 8* (2): pp. 73-93.

Krapova, Iliana. 2010. “Bulgarian Relative and Factive Clauses with an Invariant Complementizer.” *Lingua 120* : pp. 1240-1272.

Leu, Tom. 2010. “The Internal Syntax of `jeder' (`every').” *Language Variation Yearbook 2009*: pp. 153–204.

Leu, Tom. 2015. *The Architecture of Determiners*. Oxford and New York: Oxford University Press.

Lipták Anikó, *On the syntax of WH-items in Hungarian*, Doctoral dissertation, Utrecht: LOT.

Manzini, Maria R. and Leonardo M. Savoia. 2011. *Grammatical Categories: Variation in Romance Languages*. (Cambridge Studies in Linguistics 128). Cambridge: Cambridge University Press.

Melvold, Janis. 1991. “Factivity and Definiteness.” In *MIT Working Papers in Linguistics 15*, edited by Lisa Cheng and Hamida Demirdache, pp. 97-117. Cambridge: MIT press.

Mitrović, Ivana. 2012. “Relative Clauses in Serbian. ” Handout of a talk given in the *Syntax Lab*, University of Maryland, Sept 11.

Rizzi, Luigi. 1997. “The Fine Structure of the Left Periphery.” In *Elements of grammar: A Handbook of Generative Syntax*, edited by Liliane Haegeman, pp. 281-337. Dordrecht: Kluwer.

Rizzi, Luigi. 2004. “Locality and Left Periphery.” In *Structures and Beyond. The Cartography of Syntactic Structures Vol. 3*, edited by Adriana Belletti, pp. 223-251. New York NY: Oxford University Press.

Rizzi, Luigi. 2013. “Syntactic Cartography and the Syntacticisation of Scope-discourse Semantics.” In *Mind, Values and Metaphysics – Philosophical Papers Dedicated to Kevin Mulligan*, edited by Anne Reboul, pp. 517-533. Dordrecht: Springer.

Roussou, Anna. 1992. “Factive Complements and Wh-Movement in Modern Greek.” *UCL Working Papers in Linguistics 4*: pp. 123-147.

Roussou, Anna. 2010. “Selecting complementizers.” In *Exploring the left periphery*, edited by Kleanthes Grohmann and Ianthi Tsimpli. *Lingua thematic issue*, *120:* pp. 582-603.

Schlenker, Philippe. 2005. “The Lazy Frenchman's Approach to the Subjunctive”. In *Romance Languages and Linguistic Theory 2003: Selected Papers from Going Romance 2003*, edited by Twan Geerts, Ivo van Ginneken and Haike Jakobs, pp. 269-309. Amsterdam/Philadelphia: John Benjamins.

Sočanac, Tomislav. 2017. *Subjunctive Complements in Slavic Languages: A Syntax-Semantics Interface Approach*. Doctoral dissertation, University of Geneva.

Starke, Michal. 2001. *Merge Dissolves into Move*. Doctoral dissertation, University of Geneva.

Szabolcsi, Anna, James Doh Whang, and Vera Zu. 2014. "Quantifier words and their multi-functional(?) parts." *Language and Linguistics 15*: pp. 115-155.

Todorovic, Natasa. 2012. *The Subjunctive and Indicative da-Complements in Serbian A Syntactic-Semantic Approach*. Doctoral dissertation, University of Illinois at Chicago.

Tóth, Enikő. 2008. *Mood Choice in Complement Clauses.* Frankfurt: Peter Lang.

---

[1]♣ This chapter is a modified and extended version of Baunaz (2016b), both in data and analysis. I would like to thank Boban Arsenijević, Metin Bagriacik, Alexia Ioannidou, Ciro Greco, Liliane Haegeman, Illiana Krapova, Eric Lander, Teodora Radeva-Bork, Anna Roussou, Vesela Simoneva, Tomislav Sočanac, Andrew Weir, as well as the audience of the SLE 44 workshop *Nanosyntax* and NELS 46, for judgments and/or comments on previous versions of this paper. I am also very grateful to the two reviewers of this volume for various insightful comments helping improve this paper. All remaining errors are my own. This research has been partly supported by the Swiss National Foundation (grant: PA00P1_145313), as well as FWO project 2009-Odysseus-Haegeman-G091409.

[1] There is a debate concerning the status of SC *da*: either there are two homophonous items: 'declarative' *da* and modal *da*, or only one *da*. See Todorovic (2012) and references cited there for details.

[2] For syntactic complexity inside of 'complementizers', see also Leu (2015), who argues that German *dass* is two complementizer heads: *d-* and *-ass*.

[3] *Da* here is register-bound: it appears in higher/more formal registers (Boban Arsenijevic, Tomislav Sočanac, p.c.).

[4] Recall that when Serbian speakers embed *što* under 'remember'-type verbs, factivity is weakened, i.e. the matrix verb is interpreted as a relatively veridical verb (see Section 2. See

also (4) for relevant examples). When SC speakers embed *što* under 'regret'-type verbs, factivity is strengthened, i.e. the matrix verb is interpreted as a strongly veridical verb. This is why the % is inverted in Tables 6.5 and 6.6 vs. Table 6.4/6.2.

[5] The French data are based on Baunaz and Puskás (2014), Baunaz (2015). The MG/SC data are based on Baunaz (2014, 2015). The reader is refered to these papers for discussion about focus and topic extractions in French, SC, and MG, where it is shown that focus extraction behaves like *wh*-extraction, but not topic extraction.

[6] The SC *što*-variants are discussed in Section 4.4.

[7] Why (43a) is fine and (44a) ungrammatical still remains to be explained, even under a relativization analysis.

[8] See Haegeman and Ürögdi (2010ab) on operator movement and factivity, see also Aboh (2005) and Krapova (2010) about some factives being relative clauses. See also Arsenijevic (2009) and Kayne (2009) for embedded clauses being relative clauses.

[9] OpP also spells out an fseq. As this is beyond the scope of this paper, the reader is referred to Baunaz and Lander (under review) for an analysis. This means that OpP is constituted of (potentially various) features.

[10] Even though BaseP in Figure 6.2 is at the core of the internal structures of complementizers, it does not select anything, syntactically speaking: it is the higher layers that determine where our nano-structure gets inserted in the clause (see also De Clercq Chapter 7).

[11] Also relevant to this discussion is recent work by Szabolcsi, Whang & Zu (2014) and the references cited there (thanks to a reviewer for pointing this out to me).

[12] Thanks to Eric Lander for discussing this with me and for providing the relevant example.

[13] This paper only focuses on features that participate in RM, but there might be more features sandwiched between PartP and nP, like features specifying n ('thing', 'form', 'body', etc.) or

phi-features, i.e. features that would account for the differences between *que* and *qui*. This is left for future research.

[14] But see Baunaz and Lander (to appear) for development.